EFL writing in the disciplines: Insights from the VESPA learner corpus

Magali Paquot
Centre for English Corpus Linguistics
Université catholique de Louvain
Outline

- Learner corpus research
  - Learner corpora: data and methods
  - Insights for EAP
- The VESPA project
- Illustrations
  - Granger & Paquot (2009)
  - Hugon (2011)
- Conclusion and pedagogical applications
Learner corpora

- Electronic collections of authentic language data produced by learners of a foreign/second language (Granger, 2009)
- Strict design criteria in order to control the wide range of variables that affect learner language
  - learner variables: age, proficiency level, mother tongue background, etc.
  - task variables: field, genre, topic, etc.
- *International Corpus of Learner English* (Granger et al, 2009)
Contrastive Interlanguage Analysis (CIA) (Granger, 1996)

- NL: native language; IL: interlanguage
Insights of learner corpus research for EAP

- Gilquin et al., 2007; Paquot, 2010
- Academic vocabulary in learner writing
  - International Corpus of Learner English
  - Academic Keyword List (Paquot, 2010)
  - Organizational functions: cause and effect, compare and contrast, conclude, exemplify, introduce topic, etc.
A common core of interlanguage features

- Limited lexical repertoire
- Lack of register awareness
- Lexico-grammatical and phraseological specificities
- Semantic misuse (e.g. connectors and abstract nouns)
- Extensive use of chains of connective devices
- Marked preference for sentence-initial position of connectors
**Limited lexical repertoire: cause and effect**

<table>
<thead>
<tr>
<th></th>
<th><strong>overuse</strong></th>
<th><strong>no statistical difference</strong></th>
<th><strong>underuse</strong></th>
<th><strong>TOTAL</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>nouns</strong></td>
<td></td>
<td></td>
<td></td>
<td>11 [100%]</td>
</tr>
<tr>
<td>root, consequence</td>
<td>2 [18.9%]</td>
<td>4 [36.7%]</td>
<td>5 [45.4%]</td>
<td>17 [100%]</td>
</tr>
<tr>
<td>cause, reason, result</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>origin, effect, outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>verbs</strong></td>
<td></td>
<td>cause, bring about, contribute to, lead to</td>
<td>stem, provoke, result in, yield, arise, derive, emerge, follow, trigger</td>
<td>17 [100%]</td>
</tr>
<tr>
<td>cause</td>
<td>1 [5.9%]</td>
<td></td>
<td>13 [76.5%]</td>
<td></td>
</tr>
<tr>
<td>bring about, contribute to, lead to</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>adjectives</strong></td>
<td>0</td>
<td>1 [50%]</td>
<td>1 [50%]</td>
<td>2 [100%]</td>
</tr>
<tr>
<td>responsible (for)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>consequent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>adverbs</strong></td>
<td>4 [40%]</td>
<td>2 [20%]</td>
<td>4 [40%]</td>
<td>10 [100%]</td>
</tr>
<tr>
<td>as a result, as a consequence, so</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>therefore, in consequence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>accordingly, thus, hence, thereby</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>prepositions</strong></td>
<td>3 [27.3%]</td>
<td>6 [54.5%]</td>
<td>2 [18.2%]</td>
<td>11 [100%]</td>
</tr>
<tr>
<td>because of, due to, thanks to</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>as a result of, as a consequence of</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in view of, in (the) light of</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>conjunctions</strong></td>
<td>2 [40%]</td>
<td>0</td>
<td>3 [60%]</td>
<td>5 [100%]</td>
</tr>
<tr>
<td>because, this/that is why</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for, so that, on the grounds that</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>12 [21.4%]</td>
<td>16 [28.6%]</td>
<td>28 [50%]</td>
<td>56 [100%]</td>
</tr>
</tbody>
</table>
Lack of register awareness

- Overuse of more informal linguistic features
  - the verb *say*; the (sentence-final) adverb *though*; the preposition *like*
  - *I am going to talk about*; *by the way*; *of course*

- Underuse of words and phrasemes that are typical of academic discourse
  - the adverb *notably*; the verbs *illustrate* and *exemplify*
  - ... *is discussed in*; *as suggested by*...
Learner corpus data

- Predominantly collected in the context of foreign/second language courses for general purposes
  - Argumentative essays (e.g. ICLE)
    - ‘Opinion’ writing: death penalty, Europe, religion, feminism
  - Speech (e.g. interviews: LINDSEI)
    - Free discussion + picture description
LSP learner corpus (cf. Granger / Paquot, 2012)

- discipline and genre-specific texts written by learners within the framework of LSP or content courses (e.g. work by L. Flowerdew)
- Discipline-specific student writing “has tended to be collected for individual scholarly purposes rather than as part of formal corpus-building projects” (Alsop & Nesi, 2009: 75).
The VESPA project
Varieties of English for Specific Purposes dAtabase (VESPA)

- Corpus of **English for Specific Purposes texts** written by L2 writers from various mother tongue backgrounds
- L2 texts in a wide range of
  - **disciplines** (e.g. linguistics, business, medicine)
  - **genres** (e.g. papers, reports, dissertations)
  - degrees of **writer expertise** (bachelor, master).
International project

- 9 languages
  - Dutch, French, German, Italian, Norwegian, Polish, Spanish, Swedish and Turkish

- Disciplines covered so far
  - Linguistics (French, Norwegian, Spanish, German, Turkish)
  - Engineering (Spanish & Swedish)
  - Business (Polish, Italian, Turkish)
A small comparison of the Interactional and Informational Themes in the 2007 Nobel Peace Prize Lectures

1. Introduction
This paper will consider the Interactional and Informational Themes in the Nobel Lectures given by Al Gore and Rajendra Kumar Pachauri, who represents the IPCC. These two speakers have applied a very different approach to this task, and this is reflected in their Lectures for instance in the Theme choices (the texts will be referred to as the P-text and the G-text). Turning from the Interactional and Informational Themes, the discussion in this paper will turn to what the speakers have presented as new Themes. The appendix lists what this paper has analysed as new Themes.

2. Theory and method
This paper applies the framework Margaret Berry (1995) outlined in a study, where she investigated thematisation in children’s writing. The children had been assigned to write a paper for a competition where the instructions for the writing task indicated, a bit ambiguously, that they should write for a guidebook or a travel brochure. These two genres have different aims; the first requires informational writing, the second requires promotional writing. This paper will not go into the details of Berry’s study, but her framework is outlined below.

Berry (1995) considers:
Word macros

A small comparison of the Interactional and Informational Themes given by Al Gore and Rajendra Kumar Pachauri, who represents the IPCC, has applied a very different approach to this task. This is related to the Lectures for instance in the Theme choices (the texts will be referred to as G-text). Turning from the Interactional and Informational Themes, the discussion in this paper will turn to what the speakers have presented as new Themes. The appendix lists what this paper has analysed as new Themes.

1. Introduction

This paper will consider the Interactional and Informational Themes given by Al Gore and Rajendra Kumar Pachauri, who represents the IPCC. These speakers have applied a very different approach to this task, and this is related to the Lectures for instance in the Theme choices (the texts will be referred to as G-text). Turning from the Interactional and Informational Themes, the discussion in this paper will turn to what the speakers have presented as new Themes. The appendix lists what this paper has analysed as new Themes.

2. Theory and methods

This paper applies the framework described by Berry (1995) outlined in a study, where...
Insights from the VESPA learner corpus

Granger & Paquot (2009)
Hugon (2011)
**VESPA-FR-LING**

<table>
<thead>
<tr>
<th>L1</th>
<th>Number of texts</th>
<th>Number of words</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>60 (pilot: 119)</td>
<td>170,608 (359,990)</td>
</tr>
</tbody>
</table>
Granger / Paquot (2009)

- Investigate whether there is really no common core vocabulary that can be taught in a general EAP context
- Comparison of VESPA-FR-LING and native corpora
- Lexical verbs
  - Keyword analysis to select EAP-like verbs
Collocations of the verb analyze in VESPA-FR-LING

- Typical EAP collocations
  - I will explain how I have used Wordsmith software to analyze my data.
  - Of course it is not the purpose of this paper to analyze the data from all possible points of view.
  - We will present a practical part where both our corpora are analysed in terms of construction, number, semantics and inner structure.
Atypical or clumsy collocations

- It is now interesting to analyse these results more deeply by looking at the way the different speakers use these three verbs.
- Wordsmith Tools Controller enables me to analyse these data as accurately as possible.
- This paper tries to analyse some processes of derivation used by Panofsky, by focusing on adjectivalizers and prefixes combined with proper nouns.
- In this case, the grammatical properties of the item can be guessed by analyzing its stress placement or its context.
- The table below analyses this phenomenon.
“analyze” in VESPA-FR-LING: active vs. passive structures

- Overuse of active structures introduced by a first person pronoun
  - *I will analyse* the syntactic structure of the sitcom dialogs in comparison to a corpus of compiled authentic conversations.
  - *We also analysed* some properties of the root and asserted that it often was a substantive.

- Underuse of passive structures
French learners’ “pet pattern”:

Let us analyze

- **Let’s analyse** the Greek prefix anti- and **not forget** that Brinton includes it among negative prefixes whereas Lightner and Quirk et al. consider it as a prefix of opposition.

- **Let me thus analyse** the 3ADJ entries more closely.

Mastery of field-specific collocations

- **A noun phrase can be analyzed** according to both its form and its grammatical function.
- As Biber (2002) points out, linguistics can analyze both the structure and the use of language but the emphasis has always been put more on the first one than on the second one.
- Jespersen analyses after-, off-, out-, over-, under- and up- as particles.
- **The prefixes we are analysing** are very often used in nicknames.
- Some terms and concepts that could be useful to analyse our corpus will also be defined and illustrated in this section.
Hugon (2011)

- ‘Contrast’ in VESPA-FR-LING

<table>
<thead>
<tr>
<th>Part of speech</th>
<th>Lexical items</th>
</tr>
</thead>
</table>
| Nouns          | contrast, difference, distinction  
|                | *Strong opposition*: the contrary, the opposite, the reverse |
| Adjectives     | contrasting, different, differing, unlike  
|                | *Strong opposition*: contrary, opposite, reverse |
| Verbs          | contrast, differ, distinguish, differentiate |
| Adverbs        | by contrast, in contrast, by comparison, in comparison, on the other hand  
|                | *Strong opposition*: on the contrary, conversely |
| Prepositions   | unlike, in contrast to, in contrast with, in comparison with, by comparison with, as against, as opposed to, versus  
|                | *Strong opposition*: contrary to |

*Table 3.3.1: the function of contrast (MED2: IW7-IW9)*
Data and method

- VESPA-FR-LING ‘pilot corpus’
  - B3: 62 texts; 160,813 words
  - M1: 63 texts; 160,974 words
- Reference corpus: 109 scientific articles in linguistics from the Louvain Corpus of Research Articles (LOCRA); 1,017,517 words
Main results

- Items overused: difference, distinction, opposite; different; distinguish, differentiate; on the other hand, on the contrary, in comparison with, contrary to
- Items underused: contrast (n.), contrasting, contrast (v.), in contrast, in contrast to
- ‘Collocational teddy beards’: significant difference, clear distinction
- Infelicitous combinations
‘distinguish’

- to recognize and understand the differences between two or more things or people
- to be the thing that makes someone or something different or special

<table>
<thead>
<tr>
<th></th>
<th>Total VESPAL</th>
<th></th>
<th>LOCRAAL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq.</td>
<td>%</td>
<td>Freq.</td>
<td>%</td>
</tr>
<tr>
<td>Meaning 1</td>
<td>133</td>
<td>94.3</td>
<td>137</td>
<td>70.6</td>
</tr>
<tr>
<td>Meaning 2</td>
<td>8</td>
<td>5.7</td>
<td>57</td>
<td>29.4</td>
</tr>
<tr>
<td>Total</td>
<td>141</td>
<td>100</td>
<td>194</td>
<td>100</td>
</tr>
</tbody>
</table>
B3 vs. M1 students

- ‘distinction’ and ‘distinguish’ are largely overused by B3 students
- Connectors more often used in sentence-initial position in B3
- More infelicitous combinations in B3
- Erroneous use of ‘on the other side’ instead of ‘on the other hand’ only in B3
- More erroneous uses of ‘on the contrary’ in B3
- More academic-like combinations in M1: marked difference vs. big/huge difference; strongly differ vs. differ a lot
- More discipline-specific combinations in M1: lexical / linguistic / grammatical difference
Conclusion
Learner corpus research and applied linguistics

- Native corpora: frequency and typicality
- Learner corpora: degree of difficulty (overuse, underuse, error, etc.)

- Macmillan English Dictionary for Advanced Learners (2nd edition)
  - Get it right! notes
  - 12 writing sections
  - 6 lexico-grammatical sections
What is specific about teaching *vocabulary* for specific purposes?

- Two trends
  - Generalizing trend: focus on cross-disciplinary similarities (Coxhead 2000)
  - Particularizing trend: focus on field-specific differences (Hyland and Tse 2007)

- Happy medium
  - Granger and Paquot (2009)
Louvain EAP Dictionary (LEAD)

The Louvain EAP dictionary

claim (v."")

to state that something is true, although there is no clear evidence and others may disagree

- David Graeber, in his 2004 study 'Fragments of an Anarchist Anthropology', claimed that the university functions as a hierarchical disciplining device that places graduates in state and corporate bureaucracies.
- A French artist and art historian claims to have cracked the real Da Vinci code, the mystery of how Leonardo painted the Mona Lisa.

Warning

Learners sometimes use the verb claim when other verbs such as argue, show, or suggest would be more appropriate:

- The Ministry of Education claimed that children should get more free time.
- The Ministry of Education insisted that children should get more free time.

The verb claim should only be used:

- to report something that you think has not been proved:

  Echoing Plato, Galileo claimed that God had created the world according to mathematical proportions.

- or to show that you do not agree with what you are reporting:

  The party may claim to be in the vanguard of history, but its clock seems to have stopped in the 1950s.
nevertheless (adv.)

You can use the adverb **nevertheless** for introducing a concession, that is, to show that something remains true, valid, or accurate while establishing the truth of another statement in such a way as to make your first statement seem less important.

The adverb **nevertheless** is used:

- inside the sentence:

  *The raising of children is a private family matter which is **nevertheless** the subject of public debate and concern.*

- at the beginning of the sentence, followed by a comma:

  *He was evidently much more successful in the art of privacy than she had been. **Neverthelesss**, there were certain facts which he could not conceal.*

**Be careful!**
The adverb **however** is much more frequent than **nevertheless** and **nonetheless**. Only use **nevertheless** and **nonetheless** when you want to emphasize that the statement you are introducing is true, even though your previous sentence makes this seem very unlikely.

**Academic writing**

![Frequency of adverbs](image-url)

- **however**
- **nevertheless**
- **nonetheless**
**LEAD & discipline-awareness**

---

**The Louvain EAP dictionary**

<table>
<thead>
<tr>
<th>Headwords</th>
<th>function</th>
<th>possibility (n.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>expresses possibility or uncertainty</td>
<td>possibility (n.)</td>
</tr>
<tr>
<td></td>
<td>In French</td>
<td>possibility alternative</td>
</tr>
</tbody>
</table>

**possibility** (n.)

1. a chance that something could happen or be true
   - The winner of the prize for research spoke of the **possibility** of finding a cure for multiple sclerosis.

2. something that you can decide to do in a particular situation (similar to **option**)  
   - There is a **wide range of possibilities** for collaborative research with our researchers and research units.
   - Forth Road Bridge workers are considering the **possibility** of industrial action in the wake of internal changes affecting payment and working conditions.

**Warning**

Many learners use the noun **possibility** followed by to + INFINITIVE. When this noun expresses degrees of certainty, it should be used in the pattern **possibility** + of + V-ing:

- Vocational schools offer **the possibility to take** university level courses in addition to the vocational training.
- Vocational schools offer **the possibility of taking** university level courses in addition to the vocational training.

**This unique aspect of leaders' problem-solving activities raises the possibility of a host of social errors.** (BUS)

**These data raise the possibility of a type of leadership that promotes positive intergroup relations.** (BUS)

**Milliken and Martinez (1996) raised the possibility that racial diversity can have negative effects on outcomes like satisfaction with group performance.** (BUS)

---

**Collocations**

- possibility
- other, new, real, further, various, very, mere, different, obvious, only, interesting, greater, theoretical, alternatives, exciting, serious, distinct, reasonable, likely, remote, practical, realistic, numerous, following, logical

- V + possibility
  - consider, raise, offer, exclude, suggest, allow, lease, discuss, explore, deny, see, open, create, avoid, remain, recognise, reduce, provide, preclude, become, investigate, eliminate, recognize, reject, exhaust

- possibility + V
  - exist, arise, remain, suggest, seem

**Lexical bundles**

- the possibility of
- the possibility that
- another possibility is
- there is the possibility of

**More corpus examples from ...**

- business
We need more VESPA partners!

Thank you!
Hasselgård, H. (2012) *It*-clefts in English L1 and L2 academic writing


Paquot, M. (under preparation) Features of learner writing vs. novice writing in the VESPA corpus.